

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

POLICY PERSPECTIVES

Optimizing the Leveraging of Real-World Data to Improve the Development and Use of Medicines



Marc L. Berger, MD^{1,*}, Craig Lipset, BA¹, Alex Gutteridge, MS¹, Kirsten Axelsen, MS¹,
Prasun Subedi, PhD¹, David Madigan, PhD²

¹Pfizer, Inc, New York, NY, USA; ²Columbia University, New York, NY, USA

ABSTRACT

Health research, including health outcomes and comparative effectiveness research, is on the cusp of a golden era of access to digitized real-world data, catalyzed by the adoption of electronic health records and the integration of clinical and biological information with other data. This era promises more robust insights into what works in health care. Several barriers, however, will need to be addressed if the full potential of these new data are fully realized; these will involve both policy solutions and stakeholder cooperation. Although a number of these issues have been widely discussed, we focus on the one we believe is the most important—the facilitation of greater openness among public and private stakeholders to collaboration, connecting information and data sharing, with the goal of making robust and complete data accessible to all researchers. In this way, we can better understand the consequences of health care

delivery, improve the effectiveness and efficiency of health care systems, and develop advancements in health technologies. Early real-world data initiatives illustrate both potential and the need for future progress, as well as the essential role of collaboration and data sharing. Health policies critical to progress will include those that promote open source data standards, expand access to the data, increase data capture and connectivity, and facilitate communication of findings.

Keywords: big data, data access, health research, health policy, real-world data.

Copyright © 2015, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Introduction

Health research, including health outcomes and comparative effectiveness research, is on the cusp of a golden era of access to digitized real-world data that promises to transform the way in which we understand and practice medicine. Part of this transformation will be driven by the quantity of real-world data that will be generated—as well as the broader interest in “Big Data.” Real-world data are collected outside of a clinical trial and used for health care decision making [1]. Real-world data can include electronic medical records originating from health care providers, data used to coordinate and pay for care, and pharmacy data used to fill prescriptions. Data may also be collected in patient registries or pragmatic clinical trials. Internet searches and social media are also a growing source. Real-world data become “big data” when multiple data sets are combined. Definitions of big data can vary, but for this context, the term can be defined by “high-volume, high velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making” [2]. The

big data paradigm has been driven by widespread efforts to digitize and synthesize existing data sources (e.g., electronic health records [EHRs]) as well as novel mechanisms to capture both clinical and biological data (e.g., wearable health devices). Whether small and “real world” or “big,” these new data sources hold great promise to help improve our ability to develop new treatments and cures, to predict which ones are most effective, for which patients, and at what cost. For example, new types of real-world data could allow researchers to readily identify target populations of interest, and to identify patients who may be more likely to respond to treatment. In addition, these data could significantly improve health care delivery by enabling the development of a learning health care system that provides more rapid feedback to providers and patients to allow them to optimize treatment. They could allow health plans to develop benefits that are tailored to the patient and value based, varying cost sharing and access on the basis of clinical need. These new data will certainly have an impact on how we monitor both the safety and the effectiveness of treatment, and these efficiencies will likely accelerate efforts to replace the current, volume-based

* Address correspondence to: Marc L. Berger, Pfizer, Real World Data and Analytics, 235 East 42nd Street, New York, NY 10017.

E-mail: marc.berger@pfizer.com.

1098-3015/\$36.00 – see front matter Copyright © 2015, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

<http://dx.doi.org/10.1016/j.jval.2014.10.009>

fee-for-service health care system into one that allows for more efficient spend of health care dollars.

Real-World Data Trends in Health Care

Currently, there are multiple public and private efforts to digitize and aggregate health information from administrative claims, EHR, and laboratory tests. Some of these efforts are also collecting additional sources of data including genomic data, patient-reported data, and biometric data from sensors. Although combining data across sites of care and broadening access has potential value for health research, it poses a risk to privacy. Even with the protections provided through the Health Insurance Portability and Accountability Act [3], there is still the risk of reidentification, particularly if data sets are merged with other information such as voter registration. Removing personal identifiers, aggregating small samples as required through the Safe Harbor and Limited Dataset provisions of Health Insurance Portability and Accountability Act along with careful consideration of what is available through other public use data sets, may reduce that risk [4,5]. We contend that there exist or there are emerging solutions that would permit the “mashing” together of data sets at a patient level with limited risk to privacy and that the more difficult issue is that of data ownership and access.

Patient Ownership and Access of Data

Although some governments are making efforts to permit patients and researchers greater access to data, most of the health-related data are aggregated and curated by private companies. In most cases, charging for access to the data is integral to their business plans. In many cases, the data have been transformed in some fashion by the company collecting the data, and therefore it considers the data as intellectual property. Access may also be restricted to governmental-sponsored data sets. Some have argued that one way to reduce access limitations is to give patients broader rights to control the use of their own data. When patients have access to their health data, 91% of the time they are willing to share that data to benefit research [6]. Thus, ready access to data may be facilitated by the advocacy of patients.

Patients are sharing health information with one another. A pioneer in promoting the involvement of “citizen scientists” has been PatientsLikeMe. By allowing patients to publicly track health indicators, real-world data can be aggregated and analyzed in an effort to determine their best course of treatment. When an open-label study with 44 patients suggesting lithium could delay the onset of amyotrophic lateral sclerosis was published in 2008 [7], hundreds of patients in the PatientsLikeMe community began to use lithium, self-tracking symptoms and progression. Patient-LikeMe was able to use this data to publish findings in a major scientific journal, showing that lithium did not slow amyotrophic lateral sclerosis [8]. Although randomized trials were ongoing to attempt to replicate the original study, PatientsLikeMe was able to publish preliminary findings from data provided by these “citizen scientists” in only 9 months and at a very low cost.

Health care providers have been granted billions of dollars in incentives to support the implementation of EHRs through the Health Information Technology for Economic and Clinical Health Act, part of the American Recovery and Reinvestment Act of 2009. Receiving those incentives under Health Information Technology for Economic and Clinical Health Act requires providers to demonstrate Meaningful Use—a set of standards meant to ensure that EHRs bring an impact in improving the delivery of care, improving patient empowerment, and sharing information. Included in Meaningful Use is the expectation that patients will

have electronic access to their health information, and the tools to make use of that data; some have characterized this development as an important step toward “data liberation.”

One example of data liberation was launched in 2010 by the Departments of Veterans Affairs and Health & Human Services. Called the Blue Button initiative, it provides electronic access for Veterans and Medicare beneficiaries to their clinical data. By May 2012, more than 500,000 veterans had accessed their data via the Blue Button initiative and many data holders in the private sector have begun to adopt the platform.

Another move toward data liberation is the ongoing initiatives in 16 states to develop all payer claims databases, which will aggregate and collect medical, pharmacy, and dental claims data from plans including Medicaid, Health Insurance Exchange, Medicare, hospitals, and other sources to facilitate analyses by qualified researchers. States vary in their data access policies and the level of transparency on the research being conducted. Massachusetts, with a high degree of transparency and access for qualified researchers, grants access to deidentified data for research under a data use agreement. The process allows for public comment and includes representatives from a broad range of public and private stakeholders in the data release committee. Private organizations and academics have submitted applications for more than 20 studies in 2012–2013.

In contrast, the Centers for Medicare & Medicaid Services tightly manage and limit access to the national, high-quality, standardized data sets of the U.S. Medicare population. There is specifically a ban on commercial organizations using the data, which limits access to the companies that develop medicines. This represents a barrier to the data being leveraged to advance the scientific understanding of a disease for the targeting of potential new therapies and their optimal integration into health care practice. Although it is appropriate to restrict the use of this data to legitimate research designed to enhance the effectiveness of the health care system, current restrictions may distort the pool of potential researchers, the types of research that might be conducted, and, potentially, the conclusions that could be legitimately drawn from that research. One potential policy solution is to lift the ban on the commercial use of Medicare data, allowing all researchers who can maintain data privacy within specified parameters and use it appropriately for research. This could stimulate an enormous amount of understanding to optimize health care delivery for a sizeable and growing part of the American health care system.

Ability to Communicate Research

However, even if data are made available, there may still be restrictions on how the findings from analyses are communicated. The Food, Drug and Cosmetic Act’s standard for drug approval is that the manufacturer demonstrate by “substantial evidence” that the drug is safe and effective for its intended use, which generally is understood to mean data from at least two “adequate and well-controlled” clinical trials. Food and Drug Administration (FDA) regulations and enforcement practices have extended the limited statutory reach of the “substantial evidence” standard beyond the context of a drug’s approval to apply to a drug’s promotion. FDA deems safety or efficacy claims made by manufacturers in their advertising or promotional efforts that are not supported by “substantial evidence” to be “false or misleading” in violation of the Food, Drug and Cosmetic Act.

This restricts drug manufacturers’ ability to disseminate truthful, nonmisleading information about a drug’s approved uses to health care providers. For example, credible, meaningful scientific information contained in epidemiologic studies, meta-analyses, or cross-sectional surveys would not satisfy the

“substantial evidence” standard. FDA’s regulations operate selectively, prohibiting only the product’s manufacturer from discussing this data but not restricting other participants in the health care field. This creates an asymmetrical situation in which the manufacturer is not able to fully participate in open scientific discourse about its therapy. This could be addressed if Congress were to affirm that the Food, Drug and Cosmetic Act’s “substantial evidence” standard is limited to a drug’s approval. Congress could then consider requiring a more reasonable standard to assess a manufacturer’s postapproval communications, such as the Federal Trade Commission’s “competent and reliable scientific evidence” standard used to assess false and deceptive advertising. There is precedent for such action because Congress previously adopted this approach in section 114 of the Food and Drug Administration Modernization Act as applied to health care economic information intended for a knowledgeable payer audience.

Value of Public-Private Collaboration and Data Sharing

Much more will need to be done before the true value of real-world data in health care can be realized. Patient advocacy may be part of the solution, but alone will be ultimately unsuccessful for two reasons. First, there will be enormous implementation challenges that will require multisector solutions. Second, as long as business interests are not aligned, enthusiasm by data aggregators will be muted at best and progress will be slow.

One approach can be modeled on existing public-private consortia that have formed to address key noncompetitive challenges—such as the FDA Sentinel Initiative and the Observational Medical Outcomes Partnership (OMOP) to make progress on pharmacovigilance methodology. The experience of the OMOP is illustrative. Launched in 2007, this FDA-led public-private partnership focused on drug safety and obtained access to 10 large real-world data sources including medical claims and EHR. These data were deidentified and uploaded into a cloud environment. Here, the data could be interrogated within a standardized framework by a distributed network of researchers. The OMOP established a common data model, standardized vocabularies to code information, and standardized health outcome definitions. Open source tools and standardized queries were also created. The OMOP examined challenges to the validity of results from observational studies, a concern for both researchers and decision makers, and underlying ongoing controversy regarding the utility of real-world data studies in decision making. Currently, scholarly medical journals feature these studies, often with limited exploration of challenges to their validity. Thus, decision makers commonly rely on peer-review and researchers’ reputation because the ad-hoc nature of exploring bias typically eludes meaningful evaluation. The OMOP has examined thousands of epidemiologic designs for hundreds of drug-adverse outcome pairs; this has enabled it to more fully characterize bias and confidence interval coverage and has shed light on the ability of these designs to discriminate between positive and negative controls [9]. The OMOP has begun developing data-driven approaches to reduce bias and provide more accurate assessment of the robustness of associations (e.g., 95% confidence intervals that contain the true effect size 95% of the time) [10]. These results suggest that it may be possible to obtain more reliable information from real-world data studies, thereby increasing acceptance by decision makers.

Another approach is to embed health research into health care delivery as illustrated by the Salford Lung Study [11,12]. This study is evaluating Relovair in 4000 patients with chronic obstructive pulmonary disease and 5000 patients with asthma in the city of Salford, England, where the eHealth infrastructure

includes a single EHR used by both primary and secondary care. Sourcing all data from the EHR, the trial will follow patients for 1 year to monitor symptoms, exacerbations, adverse events, contacts with health care providers, and use of other therapies for symptom control. This collaboration brings together an exceptional mix of partners—pharmaceutical, health care authority, university, hospital administration, local general practitioners, and community pharmacists. Although the investment required for establishing the infrastructure for this project was substantial, it is likely that costs for similar projects in the future will fall dramatically as progress is made in the interoperability of health information technology and as new models of integrated delivery systems emerge.

Genomics, Sharing, and Collaboration

Sharing genomics data alongside the EHR produces a number of unique challenges, technical and social/ethical. The technical challenges center on the volume of data and are being met by technical improvements in data compression [13] and search technologies through initiatives such as Google Genomics [14]. The regulatory frameworks to address the social/ethical challenges are emerging through international efforts such as the Global Alliance for Genomics and Health, a coalition of more than 140 member organizations working to enable effective and responsible data sharing [15].

Conclusions

The world is changing for health research. Not everyone is ready for it. The era of digitized real-world data holds great promise in its ability to transform health care. Yet this promise can be fully realized only if access to data is broadened, if connectivity between data sets is improved, if the methods for analyzing large data sets are advanced, communication of evidence is encouraged and put in the right context, and if there are clear standards for how privacy can be maintained that also recognize that no solution is entirely secure.

A recent white paper by the 21st Century Cures initiative observed that the FDA’s review of supplemental applications for new uses or changes to a product are governed by pathways established when computers could not identify trends in statistical or clinical data anywhere close to the degree they can today, let alone what they will be capable of doing tomorrow. Considering these ongoing developments, should we be rethinking the supplemental approval processes and how real world data can be leveraged [16]? Isn’t it our collective duty to explore and rethink how we evaluate and develop medicines, talk about them, and assess health outcomes in an era of data? Don’t we owe that to patients?

Acknowledgments

We thank James Harnett, Margaret McDonald, and David Isom for their review and input on an earlier version of this article.

Source of financial support: The authors have no other financial relationships to disclose.

REFERENCES

- [1] Garrisson L, Neumann P, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force Report. *Value Health* 2007;10:326–35.

- [2] Berger ML, Doban V. Big data, advanced analytics, and the future of comparative effectiveness research. *J Comp Eff Res* 2014;3:167–76.
- [3] U.S. Department of Health and Human Services. Standards for privacy of individually identifiable health information, final rule. *Fed Reg* 2002;45 CFR, Parts 160e4.
- [4] Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. *J Am Med Inform Assoc* 2010;17:169–77.
- [5] Walter L, Sciortino S, Ranatunga D, et al. PS3-13: re-identification risk associated with sharing linked genomic and phenotypic data from the Kaiser Permanente Research Program on Genes, Environment and Health (RPGEH). *Clin Med Res* 2013;11:148.
- [6] Weitzman E, Kaci L, Mandl KD. Sharing medical information for health research: the early personal health record experience. *J Med Internet Res* 2010;12:e4.
- [7] Fornai F, Longone P, Cafaro L, et al. Lithium delays progression of amyotrophic lateral sclerosis. *PNAS* 2008;105:2052–7.
- [8] Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotech* 2011;29:411–4.
- [9] Ryan PB, Stang PE, Overhage JM, et al. A comparison of the empirical performance of methods for a risk identification system. *Drug Saf* 2013;36:143–58.
- [10] Schuemie M, Ryan P, DuMouchel W, et al. Interpreting observational studies—why empirical calibration is needed to correct p-values. *Stat Med* 2014;33:209–18.
- [11] Salford Lung Study. Available from: <http://www.salfordlungstudy.co.uk/>. [Accessed December 19, 2013].
- [12] Salford Royal NHS Trust. Groundbreaking study launched in Salford. Available from: <http://www.srft.nhs.uk/media-centre/latest-news/news-archive/news-2012/groundbreaking-study/>. [Accessed December 19, 2013].
- [13] Hsi-Yang Fritz M, Leinonen R, Cochrane G, Birney E. Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Res* 2011;21:734–40.
- [14] Google Genomics API Reference. Available from: <https://cloud.google.com/genomics/v1beta/reference/>. [Accessed October 20, 2014].
- [15] Framework for responsible sharing of genomic and health-related data. Available from: <http://genomicsandhealth.org/node/6611>. [Accessed October 20, 2014].
- [16] 21st Century Cures initiative: a call to action. Available from: <http://energycommerce.house.gov/sites/repUBLICans.energycommerce.house.gov/files/analysis/21stCenturyCures/20140501WhitePaper.pdf>. [Accessed December 19, 2013].

Erratum

In: Djawadi BM, Fahr R, Turk F. Estimating an EQ-5D value set for Malaysia using time trade-off and visual analogue scale methods. *Value Health*. 2014 December;17:814–22.

This article was incorrectly categorized as a “Patient-Reported Outcomes,” article, when it should have been included under the category, “Preference-Based Assessments.”